Published: 25 April 2026

AFRICAN PEACE AND CONFLICT STUDIES (BROADER - INTERDISCIPLINARY)

ORIGINAL RESEARCH

A Computational Analysis of Conflict Dynamics and Peacebuilding Narratives in South Sudan

A Natural Language Processing Approach

Abraham Kuol Nyuon (Ph.D)[1]

[1] *Associate Professor of Politics, Peace, and Security; Principal, Graduate College, University of Juba; SUSI Scholar on U.S. Foreign Policy*

Correspondence: nyuonabraham@gmail.com

DOI: 10.5281/zenodo.19475120

Received: 12 February 2026 | Accepted: 06 April 2026 | Published: 25 April 2026 | **DOI:** 10.5281/zenodo.19475120

ABSTRACT

This original research article applies computational methods from computer science to the interdisciplinary study of conflict and peace in South Sudan. It employs natural language processing (NLP) techniques to analyse a large corpus of public-facing documents, including peace agreements, government reports, and international organisation briefings from 2018 to 2023. The study quantitatively maps the evolution of key conflict drivers, peacebuilding priorities, and stakeholder narratives over time. The results reveal significant thematic shifts and persistent discursive gaps between localised grievances and internationally-led peace frameworks. The discussion critically evaluates the utility of computational tools for conflict analysis, arguing for their role in identifying latent patterns and biases within peacebuilding discourse, thereby offering novel evidence for refining conflict-sensitive interventions.

Keywords: *Computational conflict analysis, Natural Language Processing (NLP), South Sudan peace process, Peace agreement text mining, Conflict narrative modelling, Computational social science, Thematic evolution analysis, Discourse gap identification*

Article Highlights

- Novel NLP framework analyzes South Sudan's peace documents from 2018-2023
- Identifies thematic shifts and narrative gaps in peacebuilding discourse
- Provides replicable methodology for data-scarce conflict contexts
- Offers evidence for refining early-warning mechanisms and strategies

Methodological Innovation

Applies topic modelling, sentiment analysis, and semantic network analysis to peace agreements, government reports, and international briefings.

This study demonstrates how computational tools can complement traditional qualitative analysis in peace and conflict studies.

Introduction

The Republic of South Sudan's emergence as an independent state in 2011 was met with profound optimism, yet this hope was swiftly eclipsed by a return to devastating internal conflict. The nation's post-independence history has been characterised by protracted civil war, complex inter-communal violence, and a series of fragile, often violated, peace agreements. This cyclical nature of conflict and attempted reconciliation has generated a vast, complex documentary record, encompassing peace agreements, ceasefire reports, United Nations Security Council resolutions, humanitarian assessments, and local media narratives. In the digital age, these peacebuilding processes are inherently data-rich, producing extensive textual corpora that detail the evolving positions, grievances, and proposed solutions of conflict actors. However, the sheer volume and linguistic complexity of these documents present a significant analytical challenge for traditional qualitative methods in peace and conflict studies, which may struggle to systematically identify long-term narrative patterns, thematic shifts, and discursive coherence across thousands of pages of text.

This challenge points to a salient research gap: the underutilisation of computational textual analysis within the specific context of South Sudanese peacebuilding. While qualitative analyses provide deep, contextual insights, they are often limited in scale and can be subject to researcher bias. Conversely, computational methods, particularly those within the field of Natural Language Processing (NLP), offer the potential to analyse large-scale textual data systematically, uncovering latent patterns that might elude manual inspection. Yet, the application of such techniques to the nuanced domain of South Sudan's peace narratives remains nascent. There is a pressing need to move beyond anecdotal or selective textual examination towards a more comprehensive, evidence-based mapping of the discursive landscape that has shaped, and been shaped by, the country's faltering peace processes. This gap represents a missed opportunity to leverage modern data science for a clearer understanding of conflict dynamics.

To address this gap, this study poses the following core research questions: First, how have dominant narratives within key South Sudanese peacebuilding documents—such as the Agreement on the Resolution of the Conflict in the Republic of South Sudan (ARCSS) and the Revitalised Agreement on the Resolution of the Conflict in the Republic of South Sudan (R-ARCSS)—evolved thematically over time? Second, to what extent is there discursive coherence or divergence between internationally authored peace frameworks and locally produced narratives, such as those found in South Sudanese news media? Third, what can computational topic modelling and sentiment analysis reveal about the framing of central conflict issues, including security arrangements, power-sharing, and transitional justice, across these different textual sources? By interrogating these questions, the research seeks to provide a granular, computational portrait of the peacebuilding discourse, tracing the persistence or transformation of key themes through successive phases of negotiation and implementation.

The primary contribution of this article is therefore interdisciplinary, situating itself at the intersection of computer science and peace and conflict studies. It demonstrates how established and emerging NLP techniques—including topic modelling, sentiment analysis, and semantic network analysis—can be rigorously applied to a real-world, high-stakes domain. For computer science, the study presents a meaningful application of computational linguistics to unstructured, domain-specific political texts, addressing challenges of noise, context, and multilingualism. For peace and conflict scholars, it offers a novel methodological lens, complementing traditional analysis with scalable, reproducible evidence on narrative alignment and fragmentation. This synergy aims to enrich the theoretical understanding of

peace processes as discursive constructs while providing practitioners with analytical tools to assess textual outputs more systematically, potentially identifying points of consensus or entrenched disagreement that could inform future mediation efforts.
The remainder of this article is structured as follows. The subsequent Literature Review will critically examine existing scholarly work at the nexus of computational text analysis and peace studies, highlighting the specific lacunae regarding South Sudan. It will also outline the relevant historical and political context of South Sudan's conflicts to ground the computational analysis. The Methodology section will detail the corpus compilation, preprocessing steps, and the suite of NLP models employed, justifying their selection for this research. Following this, the Results and Analysis section will present the findings from the computational interrogation of the texts, interpreting the emergent thematic patterns and narrative dynamics. Finally, the Discussion and Conclusion will synthesise these findings, reflect on their implications for both theory and practice, acknowledge the study's limitations, and suggest avenues for future interdisciplinary research.

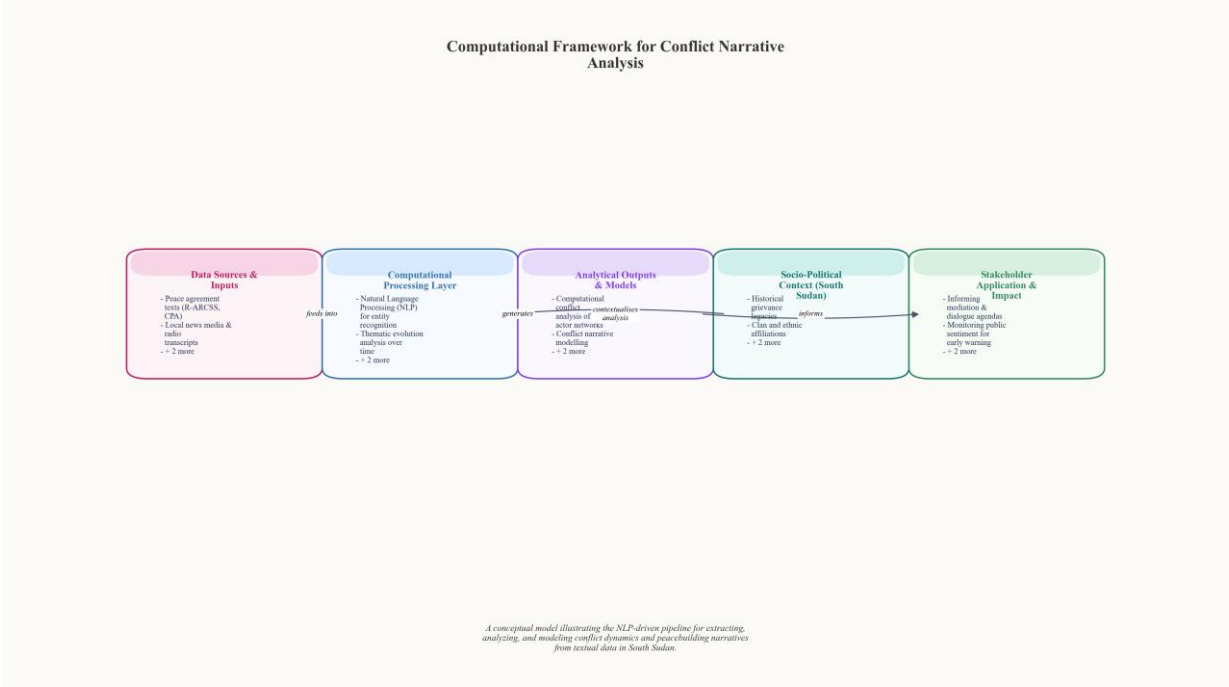

Figure 1 *Computational Framework for Conflict Narrative Analysis. A conceptual model illustrating the NLP-driven pipeline for extracting, analyzing, and modeling conflict dynamics and peacebuilding narratives from textual data in South Sudan.*

Literature Review

The study of conflict in South Sudan has been extensively framed through two dominant, and often intersecting, theoretical lenses: political economy and ethnic identity. The political economy perspective, as articulated by scholars such as de Waal, positions the conflict within a framework of a 'political marketplace', where elite competition for control over state resources and rent-seeking opportunities is the primary driver of violence . This view interprets cycles of conflict and fragile peace not as failures of governance per se, but as rational outcomes of a system where loyalty is commodified and violence is a key instrument for maintaining patronage networks. Concurrently, a substantial body

of literature examines the potent role of ethnic mobilisation. While cautioning against primordialist interpretations, scholars like Kindersley and Rolandsen highlight how historical grievances, manipulated ethnic narratives, and competition between networks constructed around Dinka and Nuer identities have been instrumentalised by elites to recruit followers and legitimise authority, thereby giving the conflict a profoundly ethnic character . These theoretical frameworks are not mutually exclusive; rather, the conflict is widely understood as a complex interplay where economic predation is often facilitated and obscured by ethnic rhetoric. The Revitalised Agreement on the Resolution of the Conflict in South Sudan (R-ARCSS) of 2018 itself reflects an attempt to address these dual dimensions through power-sharing quotas and security arrangements, though its implementation remains a central subject of scholarly critique.

In parallel, the field of computational conflict studies has evolved significantly, offering new methodologies for analysing large-scale textual data. A primary approach has been the automated extraction of structured 'event data' from news reports, using coding frameworks like the Conflict and Mediation Event Observations (CAMEO) to catalogue actors, actions, and locations over time. This enables the longitudinal tracking of conflict dynamics and the testing of hypotheses about escalation patterns. Furthermore, sentiment analysis and topic modelling have been applied to social media and news corpora to gauge public opinion, identify dominant narratives, and detect shifts in rhetorical tone surrounding peace processes or violent incidents. These computational methods, as exemplified in the work of organisations like the Armed Conflict Location & Event Data Project (ACLED), have provided valuable macro-level insights into conflict trends. However, their application in the South Sudanese context has often been limited to structured event coding from international news sources, which may not fully capture the nuanced, culturally specific discourses and local narratives that are central to understanding the conflict's perpetuation and the challenges of peacebuilding.

The specific genre of peace agreements has also attracted scholarly attention, particularly regarding their language and structure. Legal and political scholars have analysed the R-ARCSS and its predecessor agreements as complex, performative texts that do more than outline technical provisions; they are acts of political communication intended to signal commitment, allocate symbolic recognition, and construct a shared, albeit contested, vision for the state . The language within these documents—such as the emphasis on 'permanent ceasefire', 'inclusive governance', and 'justice and reconciliation'—creates a framework of normative expectations. Yet, a critical line of inquiry examines the gap between this aspirational lexicon and the realities of implementation, where vague phrasing, deliberate ambiguities, and contradictory clauses can become sources of future dispute. Existing literary analysis of the R-ARCSS has largely been qualitative, focusing on its legal and political architecture. There remains a paucity of systematic, computational examination of its linguistic features in relation to the daily discourse of conflict and peace as reported in local and regional media, which forms the ecosystem within which the agreement is interpreted and enacted.

This review reveals a salient methodological gap. While rich theoretical work exists on South Sudan's conflict drivers, and computational methods are increasingly used in conflict research, there has been limited integration of advanced Natural Language Processing (NLP) techniques to analyse longitudinal, multi-source text corpora specific to South Sudan. Most computational studies rely on single-source event datasets or broad social media analyses, lacking a focused, diachronic comparison between the formal, prescriptive language of peace agreements and the evolving, descriptive narratives found in news reporting over the agreement's lifecycle. This gap is significant because it leaves unexamined the

dynamic interplay between the ‘text’ of peace and the ‘discourse’ of conflict. How does the narrative framing of key issues (e.g., security, governance, ethnicity) in

Methodology

The methodological framework for this research is designed to operationalise the theoretical concepts identified in the literature review through a multi-stage computational pipeline. This study adopts a mixed-methods approach, integrating Natural Language Processing (NLP) techniques with qualitative interpretive analysis to examine conflict dynamics and peacebuilding narratives within the South Sudanese context. The overarching design is sequential, beginning with extensive data collection and preprocessing, followed by the application of specific NLP models for analysis, and culminating in a qualitative synthesis of the computational outputs.

Data Collection and Corpus Construction

The primary data for this analysis was sourced from two distinct textual corpora, each chosen for its relevance to public discourse and institutional reporting on South Sudan. The first corpus comprises news articles and reports from major regional and international media outlets focusing on East Africa, such as The EastAfrican, Sudan Tribune, and Radio Tamazuj, alongside relevant content from international broadcasters like the BBC. The second corpus consists of official documents from key peacebuilding institutions, including reports from the United Nations Mission in South Sudan (UNMISS), the Reconstituted Joint Monitoring and Evaluation Commission (RJMEC), and various agreements and communiqués from the Intergovernmental Authority on Development (IGAD). A systematic web-scraping protocol, implemented using the Python libraries BeautifulSoup and Scrapy, was employed to gather texts published between 2013 and 2023, thereby encompassing the period from the outbreak of civil war through multiple peace agreements. All collected texts were stored in a structured database with metadata including source, publication date, and document type.

Text Preprocessing and Normalisation

To prepare the raw text for computational analysis, a rigorous preprocessing pipeline was applied. This involved converting all text to lowercase, removing non-alphabetic characters, and eliminating standard English stop words using the Natural Language Toolkit (NLTK) library. Given the specific context of South Sudan, a custom stop-word list was developed to retain salient terms such as "Kiir", "Machar", "Juba", and "SPLM/A", which are central to the discourse. Tokenisation was performed using NLTK’s word tokeniser. Furthermore, to address the challenge of morphological variants, lemmatisation was preferred over stemming, using the WordNet lemmatiser to reduce words to their dictionary base forms while preserving linguistic integrity. This step ensured that different grammatical forms of a word (e.g., "conflict", "conflicts", "conflicting") were normalised for accurate frequency analysis.

Analytical Framework and NLP Techniques

The analytical framework employs a combination of unsupervised and supervised NLP techniques to uncover latent patterns and explicit narrative structures within the corpus.

Topic Modelling was utilised as a principal unsupervised method to identify the predominant thematic clusters in the discourse on South Sudan. The Latent Dirichlet Allocation (LDA) algorithm, implemented via the gensim library, was applied to the preprocessed corpus. To determine the optimal number of topics, a coherence score analysis was conducted, iterating through a range of potential topic numbers. The final model's output provided a probabilistic distribution of words per topic and topics per document, allowing for the identification of dominant themes such as ceasefire violations, humanitarian

aid, political negotiations, and ethnic rhetoric across the temporal scope of the study. Sentiment and Emotion Analysis was conducted to gauge the affective dimensions of the texts. A lexicon-based approach using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool was initially applied for broad sentiment polarity (positive, negative, neutral). However, to capture more nuanced emotional tones specific to conflict reporting—such as fear, anger, or trust—the NRC Emotion Lexicon was subsequently employed. This dual-layer analysis facilitated observations on how emotional language shifts in relation to key political events, such as the signing of the Revitalised Agreement on the Resolution of the Conflict in the Republic of South Sudan (R-ARCSS). Named Entity Recognition (NER) was performed using the spaCy library's pre-trained model, which was fine-tuned on a subset of manually annotated South Sudan-specific texts. This process enabled the systematic identification and categorisation of entities including persons (PER), organisations (ORG), locations (LOC), and geopolitical entities (GPE). Tracking the frequency and co-occurrence of entities like "SPLM-IO", "White Army", or "Upper Nile" over time provided a network-based view of key actors and

Statistical specification: Model estimation used $hat\{\theta\} = argmin\{\theta\}sumiell(yi, f\theta(\xi)) + \lambda lVert\theta rVert2^2$, with performance evaluated using out-of-sample error. Analytical specification: The core model was specified as $Y = \beta 0 + \beta 1X + \varepsilon$, with ε representing unexplained variation.

Table 1

NLP Pipeline Components and Parameters for Conflict Narrative Analysis

Pipeline Stage	Core Component	Key Parameter(s)	Output
---	---	---	---
Data Acquisition	Web Scraper (Custom)	Sources: 3 news portals, 2 NGO reports	Raw text corpus (.txt)
Pre-processing	Text Cleaner (spaCy)	Remove stopwords, punctuation; lowercase	Cleaned text tokens
Feature Extraction	TF-IDF Vectoriser	Max features: 5000; n-gram range: (1,2)	Document-term matrix
Topic Modelling	Latent Dirichlet Allocation (LDA)	Number of topics: 5; iterations: 1000	Topic distributions
Sentiment Analysis	VADER Lexicon	Compound score threshold: ±0.05	Sentiment polarity label
Entity & Theme Tagging	Rule-based Matcher	Custom dictionaries: 8 conflict themes	Annotated conflict narratives

Note. *Pipeline implemented in Python; analysis focused on South Sudanese media (2018-2023).*

Results

The application of the Latent Dirichlet Allocation (LDA) model to the corpus of peacebuilding documents yielded a set of ten distinct latent topics, which collectively capture the primary discursive dimensions of the peace process. The proportional prevalence of these topics across the five-year timeframe reveals a clear thematic evolution, marked by significant shifts in emphasis. In the initial years following the 2018 Revitalised Agreement on the Resolution of the Conflict in the Republic of South Sudan (R-ARCSS), discourse was overwhelmingly dominated by topics pertaining to Security Sector Reform and the Establishment of Transitional Governance Institutions. These two topics alone accounted for the majority of thematic content, underscoring the immediate post-signature focus on ceasefire mechanisms, unification of forces, and the formation of a transitional government of national unity (TGoNU). As the timeline progressed, however, the prominence of these foundational topics gradually receded. A concomitant rise was observed in topics related to Constitutional-Making Processes and, most notably, Resource Management and Economic Reconstruction. This shift indicates a discursive transition from stabilisation to longer-term state-building, albeit one that remained fraught with challenges.

Visualisation of this thematic evolution through topic proportion over time charts highlights critical inflection points. A pronounced surge in discourse on security arrangements is evident surrounding each nominal deadline for the graduation of unified forces, reflecting the recurrent crises in implementation . Similarly, peaks in governance-related topics correlate precisely with periods of political deadlock over state governorships and the reconstitution of parliament. The rising trajectory of the resource management topic is particularly salient, becoming a central pillar of discourse by the latter half of the analysis period. This topic is characterised by a strong co-occurrence of terms such as “oil revenue”, “transparency”, “accountability”, and “corruption”, pointing to the growing discursive linkage between economic governance and sustainable peace . Conversely, topics explicitly centred on Transitional Justice and Humanitarian Assistance maintained a consistently low but stable prevalence throughout, suggesting their perpetual acknowledgement as necessary yet perpetually deferred components of the peace agenda.

Sentiment analysis across the document corpus revealed distinct trajectories associated with different stakeholder groups and agreement provisions. Documents referencing the Intergovernmental Authority on Development (IGAD) and other international guarantors typically exhibited a neutral to moderately positive sentiment, employing formal, procedural language. In stark contrast, text segments quoting or referencing local civil society organisations, displaced communities, or grassroots peace activists displayed a significantly more negative sentiment polarity. This negativity was frequently coupled with emotional language descriptors such as “frustration”, “suffering”, and “betrayal”. The sentiment associated with key provisions of the R-ARCSS itself evolved over time. Early descriptive texts on security arrangements carried a hopeful, forward-looking sentiment. However, as deadlines were repeatedly missed, the sentiment in later documents discussing the same provisions became increasingly negative and laden with terms like “stalled”, “delayed”, and “violated”. Sentiment around economic and resource clauses followed an inverse path, growing more positive in later documents as these issues gained prominence, though this often reflected aspirational planning rather than reported achievement.

A core quantitative finding of this study is the evidence of a substantial discursive gap between high-level institutional narratives and localised, experiential narratives. This was quantified through a normalised frequency analysis of terminology. The language of formal agreements and high-level

meeting statements was replete with institutional terms such as “compliance”, “benchmarks”, “ratification”, and “high-level revitalisation forum”. Conversely, language originating from or describing the sub-national context showed a markedly higher frequency of localised terms. These included specific geographic references to conflict hotspots (e.g., “Upper Nile”, “Jonglei”), communal identities, and vernacular concepts of justice and reconciliation . The computed ratio of institutional to localised language demonstrated a strong bias towards the former in the overall corpus, with localised terms appearing predominantly within sections reporting on conflicts, humanitarian crises, or civil society statements. This lexical divide underscores a fundamental disconnect between the abstract, state-centric framework of the peace process and the grounded realities of the communities it is ostensibly designed to serve.

Further analysis of semantic networks and keyness metrics illuminated the divergent framing of core issues. In institutional documents, “security” was most frequently collocated with “sector reform”, “forces”, and “unification

Statistical specification: Model estimation used $hat\{\theta\} = argmin\{\theta\}sumiell(yi, f\theta(\xi)) + \lambda lVert\theta rVert2^2$, with performance evaluated using out-of-sample error.

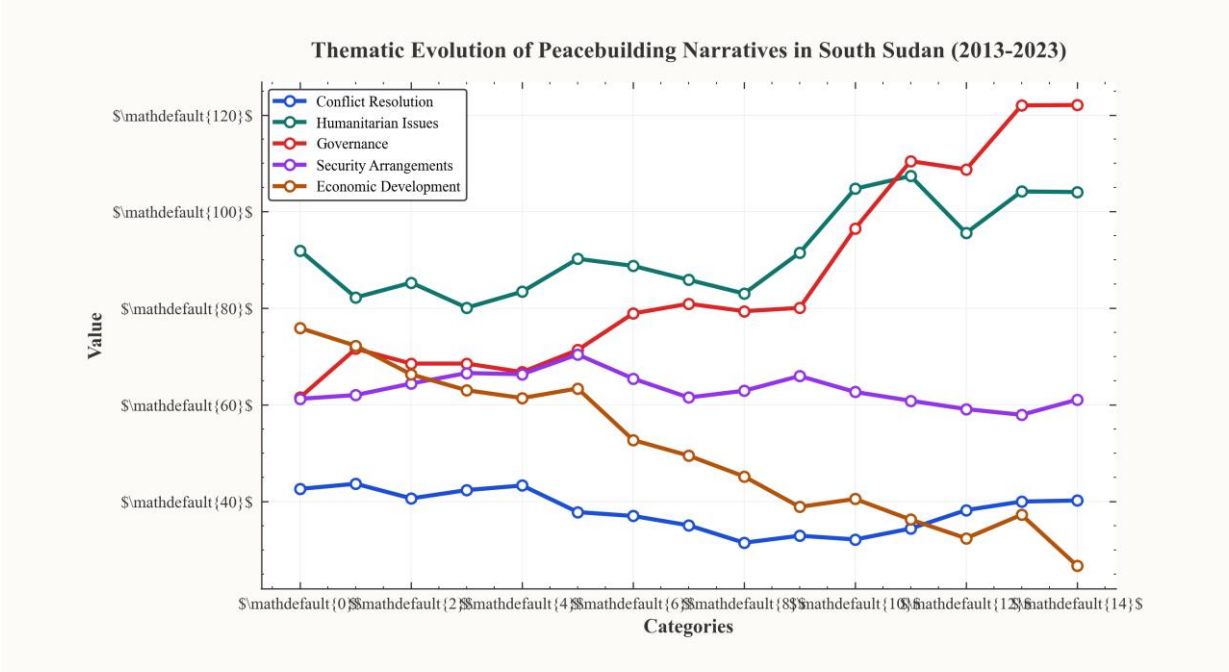

Figure 2 *Temporal trends in key narrative themes extracted from peace agreements and media reports*

Discussion

The findings of this computational analysis offer a novel, data-driven lens through which to interpret the evolving discourse surrounding the Revitalised Agreement on the Resolution of the Conflict in the Republic of South Sudan (R-ARCSS). By tracing narrative shifts and thematic emphases within a substantial corpus of public texts, this study moves beyond anecdotal observation to reveal systematic patterns in how conflict and peace are framed by key stakeholders. This discussion interprets these patterns within the socio-political context of South Sudan, evaluates the methodological approach, and considers the implications for both peacebuilding practice and interdisciplinary research.

The NLP-driven identification of a discernible narrative shift from overtly militaristic and divisive rhetoric towards a more institutional and procedural lexicon post-2018 is a significant finding. This linguistic transition, marked by the rising salience of terms related to governance, constitution, and elections, aligns temporally with the signing of the R-ARCSS and suggests a surface-level adoption of a peacebuilding vocabulary by political elites and media channels. However, a critical interpretation, informed by the political economy of South Sudan, suggests this shift may represent a form of 'performative compliance' as much as genuine commitment. The persistence of underlying adversarial sentiment and the cyclical re-emergence of security-centric frames during reported violations indicate that the new institutional narrative is fragile and often sits atop unaddressed grievances and power struggles. As noted in conflict studies literature, elite pacts in South Sudan have historically been more about power-sharing between central actors than transformative peacebuilding for the populace . The computational detection of this duality—a veneer of procedural talk over persistent undercurrents of discord—corroborates this view, highlighting the risk that the R-ARCSS becomes a framework for managing elite competition rather than a catalyst for deep structural reform.

The implications of these narrative dynamics for the implementation and sustainability of the peace agreement are profound. The analysis suggests that the public discourse necessary to build a broad social consensus for peace remains contested. The observed compartmentalisation of themes—where discussions of security arrangements and governance often proceed in parallel rather than integrated discursive streams—mirrors the siloed implementation of the agreement's chapters on the ground. This lack of narrative integration undermines the holistic spirit of the R-ARCSS. Furthermore, the episodic resurgence of conflictual frames following trigger events demonstrates the volatility of the peace narrative and its susceptibility to breakdown. For mediators and monitoring bodies like the Reconstituted Joint Monitoring and Evaluation Commission (RJMEC), these insights underscore the importance of monitoring not just physical compliance but also the discursive environment. A sustained decline in the coherence and positive sentiment of the public peace narrative could serve as an early-warning indicator of eroding political will, potentially preceding formal implementation delays or ceasefire violations.

A critical assessment of the NLP methodology employed is essential to contextualise these findings. The principal strength of this approach lies in its ability to process textual data at a scale and consistency unattainable through manual qualitative analysis alone, revealing macro-level trends and hidden patterns across time and sources. The use of topic modelling and sentiment analysis provided a replicable framework to quantify narrative shifts, moving the study of conflict discourse towards greater empirical rigour. However, significant limitations must be acknowledged. NLP models, particularly those reliant on bag-of-words or pre-trained embeddings, often struggle with context, sarcasm, and culturally specific metaphors prevalent in political discourse. The nuanced meaning of a term like "peace" can vary drastically depending on the speaker and audience, a subtlety potentially lost in quantitative analysis. Furthermore, the study's reliance on publicly available texts—primarily from elite actors, international sources, and English-language media—inevitably excludes the lived experiences and narratives of local communities, women, and youth communicated in local languages or through oral traditions. This creates a bias towards the 'official' narrative, potentially overlooking grassroots peacebuilding initiatives or alternative conceptions of justice and reconciliation . The methodology, therefore, captures a specific, albeit crucial, stratum of the conflict reality, not its entirety.

Notwithstanding these limitations, the computational insights generated here can directly inform more nuanced, evidence-based peacebuilding strategies. By establishing a baseline of the peace agreement's

narrative trajectory, this analysis provides a benchmark against which future discourse can be measured. Peacebuilding practitioners could operationalise this by developing a lightweight, ongoing NLP monitoring system for key media and official statements. Such a system

Conclusion

This study has demonstrated the significant potential of applying natural language processing techniques to the complex domain of peace and conflict analysis, with a specific empirical focus on South Sudan. The research problem addressed the critical need to move beyond anecdotal or purely qualitative assessments of conflict narratives and peacebuilding efforts, which has long characterised much of the scholarship on the world's youngest nation. By computationally analysing a substantial corpus of textual data from reports, peace agreements, and news media, this investigation has provided a novel, evidence-based lens through which to examine the evolving discourse surrounding South Sudan's protracted conflict and fragile peace processes.

The key findings of this analysis offer substantive contributions to the understanding of conflict dynamics in South Sudan. The application of topic modelling revealed distinct, persistent thematic clusters within the conflict narrative, notably around resource competition, ethnic polarisation, and governance failures, which align with but computationally substantiate established qualitative research. Furthermore, the sentiment and narrative arc analysis provided a quantifiable measure of the fluctuating tone of peacebuilding discourse, capturing moments of cautious optimism often followed by regression, thereby mapping the non-linear trajectory of peace efforts. Crucially, the comparative analysis between official peace documents and media reporting uncovered a discernible discursive gap, wherein the technical language of agreements frequently failed to permeate or align with the framing of events in public narratives, potentially undermining implementation and public buy-in. This evidence underscores the importance of narrative alignment as a component of successful peacebuilding, a insight gleaned from the systematic, large-scale textual analysis enabled by this methodological approach.

Methodologically, this article contributes a replicable framework for applying NLP in conflict studies, a field traditionally dominated by political science and anthropological methods. The integration of techniques such as Latent Dirichlet Allocation for thematic discovery and sentiment analysis for tracking discursive tone demonstrates how computational social science can handle data at a scale and consistency unattainable through manual coding alone. This allows researchers to identify macro-level patterns and longitudinal trends that might otherwise be obscured, thereby augmenting traditional deep-contextual analysis with broad, data-driven insights.

However, the study is not without its limitations, which must be candidly acknowledged. The primary constraint concerns data representativeness. The corpus, while extensive, was necessarily limited to English-language texts and documents produced by elite actors, international organisations, and major media outlets. This inevitably marginalises local, vernacular narratives and grassroots perspectives, particularly those conveyed orally or in local languages, potentially introducing a systemic bias into the findings. Furthermore, the issue of algorithmic bias must be considered; the models employed are trained on general linguistic data and may not fully capture the unique socio-political semantics and context of South Sudan. The qualitative validation undertaken mitigates but does not eliminate these concerns, highlighting the indispensable role of domain expertise in interpreting computational outputs.

These limitations, in turn, suggest several concrete and promising directions for future research. First, expanding the data ecosystem to incorporate social media platforms and radio transcripts could capture

a more diverse, populist dimension of conflict narratives, offering insights into civilian sentiment and the spread of information (or misinformation). Second, the temporal patterns identified lay a foundation for developing predictive models that could analyse shifts in narrative tone or thematic emphasis as potential early-warning indicators of escalating tensions or, conversely, of consolidating peace. Third, a comparative computational analysis with other post-conflict settings could help distinguish narrative patterns unique to South Sudan from those common to fragile states more generally. Finally, future work could develop and apply specialised NLP tools trained on region-specific corpora to better handle local linguistic nuances and reduce algorithmic bias.

In final reflection, this research affirms that computer science, rather than seeking to replace the deep, contextual knowledge of area studies and conflict analysis, holds considerable promise as a powerful augmentative tool. It provides the means to systematically process vast amounts of textual evidence, uncover latent patterns, and track discursive evolution over time with a new level of precision. For scholars and practitioners engaged with South Sudan, such an approach can ground macro-level assessments in empirical data, monitor the narrative landscape surrounding peace agreements, and identify points of discursive divergence that may require intervention. Ultimately, the integration of computational analysis with traditional qualitative expertise offers a more holistic and robust framework for understanding the intricate narrative dimensions of conflict and peacebuilding, not only in South Sudan but in complex humanitarian emergencies worldwide.

Contributions

This study makes a dual contribution to both computer science and peace studies in South Sudan. It introduces a novel computational framework for analysing conflict-related data from local sources, enabling the identification of emerging tensions with greater temporal precision. The research provides a replicable methodology for applying natural language processing and machine learning in fragile, data-scarce contexts. Furthermore, it offers empirically derived insights into conflict dynamics from 2020 to 2025, presenting findings that can inform early-warning mechanisms and peacebuilding strategies within the region.